

## **RAW SEQUENCE LISTING** **ERROR REPORT**



The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following computer readable form:

Application Serial Number: 09/849,980

Source: OIPE

Date Processed by STIC: 6/5/2001

**THE ATTACHED PRINTOUT EXPLAINS DETECTED ERRORS.**

**PLEASE FORWARD THIS INFORMATION TO THE APPLICANT BY EITHER:**

- 1) INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE APPLICANT, WITH A NOTICE TO COMPLY or,**
- 2) TELEPHONING APPLICANT AND FAXING A COPY OF THIS PRINTOUT, WITH A NOTICE TO COMPLY**

**FOR CRF SUBMISSION QUESTIONS, PLEASE CONTACT MARK SPENCER, 703-308-4212.**

**FOR SEQUENCE RULES INTERPRETATION, PLEASE CONTACT ROBERT WAX, 703-308-4216.**

**PATENTIN 2.1 e-mail help: [patin21help@uspto.gov](mailto:patin21help@uspto.gov) or phone 703-306-4119 (R. Wax)**

**PATENTIN 3.0 e-mail help: [patin3help@uspto.gov](mailto:patin3help@uspto.gov) or phone 703-306-4119 (R. Wax)**

**TO REDUCE ERRORED SEQUENCE LISTINGS, PLEASE USE THE CHECKER VERSION 3.0 PROGRAM, ACCESSIBLE THROUGH THE U.S. PATENT AND TRADEMARK OFFICE WEBSITE. SEE BELOW:**

### **Checker Version 3.0**

The Checker Version 3.0 application is a state-of-the-art Windows based software program employing a logical and intuitive user-interface to check whether a sequence listing is in compliance with format and content rules. Checker Version 3.0 works for sequence listings generated for the original version of 37 CFR §§1.821 – 1.825 effective October 1, 1990 (old rules) and the revised version (new rules) effective July 1, 1998 as well as World Intellectual Property Organization (WIPO) Standard ST.25.

Checker Version 3.0 replaces the previous DOS-based version of Checker, and is Y2K-compliant. Checker allows public users to check sequence listings in Computer Readable form (CRF) before submitting them to the United States Patent and Trademark Office (USPTO).

Use of Checker prior to filing the sequence listing is expected to result in fewer errored sequence listings, thus saving time and money.

**Checker Version 3.0 can be down loaded from the USPTO website at the following address:**

**<http://www.uspto.gov/web/offices/pac/checker>**

## Raw Sequence Listing Error Summary

### ERROR DETECTED      SUGGESTED CORRECTION

SERIAL NUMBER: 09/849,980

ATTN: NEW RULES CASES: PLEASE DISREGARD ENGLISH "ALPHA" HEADERS, WHICH WERE INSERTED BY PTO SOFTWARE

- 1 ☒ **Wrapped Nucleics  
Wrapped Aminos**      The number/text at the end of each line "wrapped" down to the next line. This may occur if your file was retrieved in a word processor after creating it. Please adjust your right margin to .3; this will prevent "wrapping."
  
- 2 ☐ **Invalid Line Length**      The rules require that a line not exceed 72 characters in length. This includes white spaces.
  
- 3 ☒ **Misaligned Amino  
Numbering**      The numbering under each 5<sup>th</sup> amino acid is misaligned. Do not use tab codes between numbers; use space characters, instead.
  
- 4 ☒ **Non-ASCII**      The submitted file was not saved in ASCII(DOS) text, as required by the Sequence Rules. Please ensure your subsequent submission is saved in ASCII text.
  
- 5 ☐ **Variable Length**      Sequence(s) \_\_\_\_\_ contain n's or Xaa's representing more than one residue. Per Sequence Rules, each n or Xaa can only represent a single residue. Please present the maximum number of each residue having variable length and indicate in the <220>-<223> section that some may be missing.
  
- 6 ☐ **PatentIn 2.0  
"bug"**      A "bug" in PatentIn version 2.0 has caused the <220>-<223> section to be missing from amino acid sequences(s) \_\_\_\_\_. Normally, PatentIn would automatically generate this section from the previously coded nucleic acid sequence. Please manually copy the relevant <220>-<223> section to the subsequent amino acid sequence. This applies to the mandatory <220>-<223> sections for Artificial or Unknown sequences.
  
- 7 ☐ **Skipped Sequences  
(OLD RULES)**      Sequence(s) \_\_\_\_\_ missing. If intentional, please insert the following lines for each skipped sequence:  
 (2) INFORMATION FOR SEQ ID NO:X: (insert SEQ ID NO where "X" is shown)  
 (i) SEQUENCE CHARACTERISTICS: (Do not insert any subheadings under this heading)  
 (xi) SEQUENCE DESCRIPTION:SEQ ID NO:X: (insert SEQ ID NO where "X" is shown)  
 This sequence is intentionally skipped  
  
 Please also adjust the "(ii) NUMBER OF SEQUENCES:" response to include the skipped sequences.
  
- 8 ☐ **Skipped Sequences  
(NEW RULES)**      Sequence(s) \_\_\_\_\_ missing. If intentional, please insert the following lines for each skipped sequence.  
 <210> sequence id number  
 <400> sequence id number  
 000
  
- 9 ☐ **Use of n's or Xaa's  
(NEW RULES)**      Use of n's and/or Xaa's have been detected in the Sequence Listing.  
 Per 1.823 of Sequence Rules, use of <220>-<223> is MANDATORY if n's or Xaa's are present.  
 In <220> to <223> section, please explain location of n or Xaa, and which residue n or Xaa represents.
  
- 10 ☐ **Invalid <213>  
Response**      Per 1.823 of Sequence Rules, the only valid <213> responses are: Unknown, Artificial Sequence, or scientific name (Genus/species). <220>-<223> section is required when <213> response is Unknown or is Artificial Sequence
  
- 11 ☐ **Use of <220>**      Sequence(s) \_\_\_\_\_ missing the <220> "Feature" and associated numeric identifiers and responses.  
 Use of <220> to <223> is MANDATORY if <213> "Organism" response is "Artificial Sequence" or "Unknown." Please explain source of genetic material in <220> to <223> section.  
 (See "Federal Register," 06/01/1998, Vol. 63, No. 104, pp. 29631-32) (Sec. 1.823 of Sequence Rules)
  
- 12 ☐ **PatentIn 2.0  
"bug"**      Please do not use "Copy to Disk" function of PatentIn version 2.0. This causes a corrupted file, resulting in missing mandatory numeric identifiers and responses (as indicated on raw sequence listing). Instead, please use "File Manager" or any other manual means to copy file to floppy disk.

09/849,980

## sequence

[illegible][illegible]

æ Ü SEQUENCE LISTING (1) General Information (i) APPLICANT: SANTEN PHARMACEUTICAL CO., LTD. (ii) TITLE OF INVENTION: Novel Polypeptide Having Water Channel Activity and DNA sequence (iii) NUMBER OF SEQUENCES: 2 (iv) CORRESPONDENCE ADDRESS: (A) ADDRESSEE: SANTEN PHARMACEUTICAL CO., LTD. (B) STREET: 9-19 Shimoshinjo 3-chome Higashiyodogawa-Ku (C) CITY: Osaka (D) STATE: Osaka (E) COUNTRY: JAPAN (F) ZIP: 533-0021 (v) COMPUTER READABLE FORM: (A) MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB, storage (B) C  
 COMPUTER: IBM PS/2 or compatibles (C) OPERATING SYSTEM: WINDOWS 95/97 (D) SOFTWARE: Microsoft Word 97 (vi) CURRENT APPLICATION DATE: (A) APPLICATION NUMBER: 09/381,810 (B) FILING DATE: 19-OCT-1999 (C) CLASSIFICATION: 435 (vii) PRIOR APPLICATION DATE: (A) APPLICATION NUMBER: JP 09 094845 (B) FILING DATE: 28-MAR-1997 (viii) ATTORNEY/AGENT INFORMATION: (A) NAME: Burton A. Amernick (B) REGISTRATION NUMBER: 24852 (C) REFERENCE/DOCKET NUMBER: 1581/00156 (ix) TELECOMMUNICATION INFORMATION: (A) TELEPHONE: (202)331-7111 (B) FAX: (202)293-6229 (2) INFORMATION FOR SEQ ID NO: 1: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 42 amino acids (B) TYPE: amino acid (D) TOPOLOGY: linear (ii) MORE CULE TYPE: peptide (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1: Met Val Gln Ala Ser Gly His Arg Arg Ser Thr Arg Gly Ser Lys Met  
 5 10 15 Val Ser Trp Ser Val Ile Ala Lys Ile Gln Glu Ile Leu Gln Arg Lys  
 25 30 Met Val Arg Glu Phe Leu Ala Glu Phe Met Ser Thr Tyr Val Met Met  
 40 45 Val Phe Gly Leu Gly Ser Val Ala His Met Val Leu Asn Lys Lys Tyr  
 50 55 60

- 1) See item 4 on Error Summary Sheet = many format errors in submitted file
- 2) See items 1 and 3 on Error Summary Sheet
- 3) See sample sequence Listing (attached) for valid format

(3) Computer: Apple Macintosh;  
 (i) Operating System: Macintosh;  
 (ii) Macintosh File Type: text with line termination  
 (iii) Line Terminator: Pre-defined by text type file;  
 (iv) Pagination: Pre-defined by text type file;  
 (v) End-of-file: Pre-defined by text type file;  
 (vi) Media: (A) Diskette—3.50 Inch, 400 Kb storage;  
 (B) Diskette—3.50 Inch, 800 Kb storage;  
 (C) Diskette—3.50 Inch, 1.4 Mb storage;  
 (vii) Print Command: Use PRINT command from any Macintosh Application that processes text files, such as MacWrite or Teach Text;  
 (4) Magnetic tape: 0.5 Inch, up to 2400 feet;  
 (i) Density: 1600 or 6250 bits per inch, 9 track;  
 (ii) Format: raw, unblocked;  
 (iii) Line Terminator: ASCII Carriage Return plus optional ASCII Line Feed;  
 (iv) Pagination: ASCII Form Feed or Series of Line Terminators;  
 (v) Print Command (Unix shell version given here as sample response—ml/dev/rmt0; lpr/dev/rmt0);  
 (g) Computer readable forms that are submitted to the Office will not be returned to the applicant.  
 (h) All computer readable forms shall have a label permanently affixed thereto on which has been hand printed or typed, a description of the format of the computer readable form as well as the name of the applicant, the title of the invention, the date on which the data were recorded on the computer readable form and the name and type of computer and operating system which generated the files on the computer readable form. If all of this information cannot be printed on a label affixed to the computer readable form, by reason of size or otherwise, the label shall include the name of the applicant and the title of the invention and a reference number, and the additional information may be provided on a container for the computer readable form with the name of the applicant, the title of the invention, the reference number and the additional information affixed to the container. If the computer readable form is submitted after the date of filing

under 35 U.S.C. 111, after the date of entry in the national stage under 35 U.S.C. 371 or after the time of filing in the United States Receiving Office, an international application under the PCT, the labels mentioned herein must also include the date of the application and the application number, including series code and serial number.

#### § 1.825 Amendments to or replacement of sequence listing and computer readable copy thereof.

(a) Any amendment to the paper copy of the "Sequence Listing" (§ 1.821(c)) must be made by the submission of substitute sheets. Amendments must be accompanied by a statement that indicates support for the amendment in the application, as filed, and a statement that the substitute sheets include no new matter. Such a statement must be a verified statement if made by a person not registered to practice before the Office.

(b) Any amendment to the paper copy of the "Sequence Listing," in accordance with paragraph (a) of this section, must be accompanied by a substitute copy of the computer readable form (§ 1.821(e)) including all previously submitted data with the amendment incorporated therein, accompanied by a statement that the copy in computer readable form is the same as the substitute copy of the "Sequence Listing." Such a statement must be a verified statement if made by a person not registered to practice before the Office.

(c) Any appropriate amendments to the "Sequence Listing" in a patent, e.g., by reason of reissue or certificate of correction, must comply with the requirements of paragraphs (a) and (b) of this section.

(d) If, upon receipt, the computer readable form is found to be damaged or unreadable, applicant must provide, within such time as set by the Commissioner, a substitute copy of the data in computer readable form accompanied by a statement that the substitute data is identical to that originally filed. Such a statement must be a verified statement if made by a person not registered to practice before the Office.

#### Appendix A—Sample Sequence Listing

##### (1) GENERAL INFORMATION:

(i) APPLICANT: Doe, Joan X. Doe, John Q.  
 (ii) TITLE OF INVENTION: Isolation and Characterization of a Gene Encoding a Protease from *Paramecium* sp.  
 (iii) NUMBER OF SEQUENCES: 2  
 (iv) CORRESPONDENCE ADDRESS:  
 (A) ADDRESSEE: Smith and Jones  
 (B) STREET: 123 Main Street  
 (C) CITY: Smalltown  
 (D) STATE: Anystate  
 (E) COUNTRY: USA  
 (F) ZIP: 12345  
 (v) COMPUTER READABLE FORM:  
 (A) MEDIUM TYPE: Diskette, 3.50 Inch, 800 Kb storage  
 (B) COMPUTER: Apple Macintosh  
 (C) OPERATING SYSTEM: McIntosh 5.0  
 (D) SOFTWARE: MacWrite  
 (vi) CURRENT APPLICATION DATA:  
 (A) APPLICATION NUMBER: 09/899,999  
 (B) FILING DATE: 28-FEB-1989  
 (C) CLASSIFICATION: 999/99  
 (vii) PRIOR APPLICATION DATA:  
 (A) APPLICATION NUMBER: PCT/US88/99999  
 (B) FILING DATE: 01-MAR-1988  
 (viii) ATTORNEY/AGENT INFORMATION:  
 (A) NAME: Smith, John A.  
 (B) REGISTRATION NUMBER: 00001  
 (C) REFERENCE/DOCKET NUMBER: 01-0001  
 (ix) TELECOMMUNICATION INFORMATION:  
 (A) TELEPHONE: (909) 999-0001  
 (B) TELEFAX: (909) 999-0002  
 (2) INFORMATION FOR SEQ ID NO: 1:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 954 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear  
 (ii) MOLECULE TYPE: genomic DNA  
 (iii) HYPOTHETICAL: yes  
 (iv) ANTI-SENSE: no  
 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: *Paramecium* sp.  
 (C) INDIVIDUAL/ISOLATE: XYZ2  
 (G) CELL TYPE: unicellular organism  
 (vii) IMMEDIATE SOURCE:  
 (A) LIBRARY: genomic  
 (B) CLONE: Para-XYZ2/36  
 (x) PUBLICATION INFORMATION:  
 (A) AUTHORS: Doe, Joan X. Doe, John Q.  
 (B) TITLE: Isolation and Characterization of a Gene Encoding a Protease from *Paramecium* sp.  
 (C) JOURNAL: Fictional Genes  
 (D) VOLUME: 1  
 (E) ISSUE: 1  
 (F) PAGES: 1-20  
 (G) DATE: 02-MAR-1988  
 (K) RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 954

BILLING CODE 3510-16-M

*Consult this*

(x1)-SEQUENCE DESCRIPTION: SEQ ID-NO:1:

ATCGGGATAG TACTGGTCAA GACCGGTGGA CACCGGTAA CCCGGTTAA GTACCGGTTA 60  
TAGGCCATTT CAGGCCAAAT GTGCCCAACT ACGCCAATTG TTTTGCCAAC GGGCAACGTT 120  
ACGTTTCGTAC GCACGTATGT ACCTAGGTAC TTACGGACGT GACTACGGAC ACTTCOGTAC 180  
GTACGTACGT TTACGTACCC ATCCCAACGT AACCACAGTG TGGTCGCAGT GTCCCAGTGT 240  
ACACAGACTG CCAGACATTC TTCACAGACA CCCC ATG ACA CCA CCT GAA CGT CTC 295  
Met Thr Pro Pro Glu Arg Leu  
-30  
TTC CTC CCA AGG GTG TGT GGC ACC ACC CTA CAC CTC CTC CTT CTG GGG 343  
Phe Leu Pro Arg Val Cys Gly Thr Thr Leu His Leu Leu Leu Gly  
-25 -20 -15  
CTG CTG CTG GTT CTG CTG CCT GGG GCC CAT GTGAGGCAGC AGGAGAATGG 393  
Leu Leu Leu Val Leu Leu Pro Gly Ala His  
-10 -5  
GGTGGCTCAG CCAAACCTTG AGCCCTAGAG CCCCCCTCAA CTCTGTTCTC CTAG GGG 450  
Gly  
CTC ATG CAT CTT GCC CAC AGC AAC CTC AAA CCT GCT GCT CAC CTC ATT 498  
Leu Met His Leu Ala His Ser Asn Leu Lys Pro Ala Ala His Leu Ile  
1 5 10 15  
GTAAACATCC ACCTGACCTC CCAGACATGT CCCCACCAGC TCTCCTCCTA CCCCTGCCTC 558  
AGGAACCCAA GCATCCACCC CTCTCCCCCA ACTTCCCCCA CGCTAAAAAA AACAGAGGGA 618  
GCCCACTCCT ATGCCTCCCC CTGCCATCCC CCAGGAATC AGTTGTTTCAG TGCCCACTTC 678  
TAC CCC AGC AAG CAG AAC TCA CTG CTC TGG AGA GCA AAC ACG GAC CGT 726  
Tyr Pro Ser Lys Gln Asn Ser Leu Leu Trp Arg Ala Asn Thr Asp Arg  
20 25 30  
GCC TTC CTC CAG GAT GGT TTC TCC TTG AGC AAC AAT TCT CTC CTG GTC 774  
Ala Phe Leu Gln Asp Gly Phe Ser Leu Ser Asn Asn Ser Leu Leu Val  
35 40 45  
TAGAAAAAAT AATTGATTTC AAGACCTTCT CCCCATTCTG CCTCCATTCT GACCATTTC 834  
GGGGTCGTCA CCACCTCTCC TTTGGCCATT CCAACAGCTC AAGTCTTCCC TGATCAAGTC 894  
ACCGGAGCTT TCAAAGAAGG AATTCTAGGC ATCCCAGGGG ACCCACACCT CCCTGAACCA 954

## (2) INFORMATION FOR SEQ ID NO: 2:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 82 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (ix) FEATURE:

(A) NAME/KEY: signal sequence

(B) LOCATION: -34 to:-1

(C) IDENTIFICATION METHOD: similarity  
to other signal sequences, hydrophobic(D) OTHER INFORMATION: expresses  
protease

## (x) PUBLICATION INFORMATION:

(A) AUTHORS: Doe, Joan X. Doe, John Q.

(B) TITLE: Isolation and Characterization  
of a Gene Encoding a Protease from  
*Paramecium* sp.

(C) JOURNAL: Fictional Genes

(D) VOLUME: 1

(E) ISSUE: 1

(F) PAGES: 1-20

(G) DATE: 02-MAR-1988

(K) RELEVANT RESIDUES IN SEQ ID NO:

2: FROM -34 TO 48

BILLING CODE 5010-10-M

Here's where sequence 2 starts (after  
the sequence data of SEQ ID NO:1:)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Thr Pro Pro Glu Arg Leu Phe Leu Pro Arg Val Cys Gly Thr Thr  
-30 -25 -20

Leu His Leu Leu Leu Leu Gly Leu Leu Leu Val Leu Leu Pro Gly Ala  
-15 -10 -5

His Gly Leu Met His Leu Ala His Ser Asn Leu Lys Pro Ala Ala His  
1 5 10

Leu Ile Tyr Pro Ser Lys Gln Asn Ser Leu Leu Trp Arg Ala Asn Thr  
15 20 25 30

Asp Arg Ala Phe Leu Gln Asp Gly Phe Ser Leu Ser Asn Asn Ser Leu  
35 40 45

Leu Val

BILLING CODE 3510-16-C